



snowflake

tutorialspoint

SIMPLY EASY LEARNING

www.tutorialspoint.com



<https://www.facebook.com/tutorialspointindia>



<https://twitter.com/tutorialspoint>

About the Tutorial

Snowflake is a cloud data platform for data storage and analytics purpose. This is a brief tutorial that introduces the readers to the basic features and usage of Snowflake. The tutorial will guide the users on what Snowflake is and how to utilize the tool for storing and analyzing the data.

Audience

This tutorial has been prepared for beginners to help them understand what Snowflake is, what all features it has, and how to use the tool.

Prerequisites

You should have a basic understanding of SQL, Database concepts, knowledge of database schema, and general caching concepts.

Disclaimer & Copyright

© Copyright 2022 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com.

Table of Contents

About the Tutorial	1
Audience	1
Prerequisites	1
Disclaimer & Copyright	1
Table of Contents	2
1. SNOWFLAKE – INTRODUCTION	5
Why Use Snowflake?	5
2. SNOWFLAKE – DATA ARCHITECTURE	7
3. SNOWFLAKE – FUNCTIONAL ARCHITECTURE	10
4. SNOWFLAKE – HOW TO ACCESS	12
5. SNOWFLAKE – EDITIONS	15
6. SNOWFLAKE – PRICING MODEL	17
What is Snowflake Credit?	17
7. SNOWFLAKE – OBJECTS	19
Snowflake Objects	19
Important Points Related to Snowflake Objects	19
8. SNOWFLAKE – TABLE AND VIEW TYPES	21
Table Types	21
View Types	22

9.	SNOWFLAKE – LOGIN.....	24
10.	SNOWFLAKE – WAREHOUSE.....	26
	Working on Warehouses using Snowflake's UI	26
	Working on Warehouses using Snowflake's SQL Interface.....	34
11.	SNOWFLAKE – DATABASE	37
	Working with Databases using Snowflake's UI	37
	Working on Databases using Snowflake's SQL Interface	41
12.	SNOWFLAKE – SCHEMA	43
	Working with Schema using Snowflake's UI	43
	Working on Schemas Using Snowflake's SQL Interface	49
13.	SNOWFLAKE – TABLE & COLUMNS	50
	Working with Tables and Columns using Snowflake's UI	50
	Working on Tables and Columns using Snowflake's SQL Interface	55
	Setup Database, Warehouse and Schema in SQL.....	56
14.	SNOWFLAKE –LOAD DATA FROM FILES.....	58
	Load Data into Tables and Columns using Snowflake's UI.....	58
	Load Data into Tables and Columns using SQL.....	63
15.	SNOWFLAKE – SAMPLE USEFUL QUERIES	66
16.	SNOWFLAKE – MONITOR USAGE AND STORAGE	72
	Checking Storage	72
	History.....	73

Monitoring	75
17. SNOWFLAKE – CACHE	78
Metadata Caching	78
Query Result Caching.....	80
Data Cache	84
18. SNOWFLAKE – UNLOAD DATA FROM SNOWFLAKE TO LOCAL.....	87
Unload Data into local file using Snowflake's UI	87
Unload Data into Tables and Columns using SQL.....	89
19. SNOWFLAKE – EXTERNAL DATA LOADING (FROM AWS S3).....	93
Using UI	93
Using SQL	96
Using UI	96
Using SQL	97
20. SNOWFLAKE – EXTERNAL DATA UNLOADING (INTO AWS S3)	99
Using UI	99
Using SQL	101
Using UI	102
Using SQL	103

1. Snowflake – Introduction

Snowflake is a cloud-based advanced data platform system, provided as Software-as-a-Service (SaaS). Snowflake provides features of data storage from AWS S3, Azure, Google Cloud, processing complex queries and different analytic solutions. The analytic solutions provided by Snowflake are faster, easy to use and more flexible than traditional databases and their analytics features. Snowflake stores and provide data near time not in actual real time.

Snowflake is advanced solution for OLAP (Online Analytical Processing) technology. OLAP is also known as online data retrieving and data analysis system using historical data. It processes complex and aggregated queries with low number of transactions. For Ex: Getting number of orders, sales amount in last month for a company, number of new users list in the company in last quarter etc. Snowflake is not used as OLTP (Online Transactional Processing) database. OLTP database usually contains real time data with a high volume of small data transactions. For Ex: Inserting customer's order detail, register a new customer, tracking order delivery status etc.

Why Use Snowflake?

Snowflake provides Data Platform as a Cloud Service.

- There is no hardware neither virtual nor physical to select, install, configure or manage from client side.
- There is no software to install, configure or manage to access it.
- All ongoing maintenance, management, upgrades and patching are owned by Snowflake itself.

Traditional databases for analytics solutions are complex in architecture, costly and constrained while Snowflake is rich in concept of Data Engineering, Data Lake concept, data warehouse, Data Science, Data Application and Data Exchange or sharing. It is easy to access and use without having constraint of data size and storage capacity. User must administrate only their data; all data platform related administrations are done by Snowflake itself.

Apart of these, Snowflake also has the following features:

- Simple, reliable data pipelines in multi languages like Java, Python, PHP, Spark, Ruby etc.
- Secured access, very good performance and security of data lake.
- Zero administration for tool, data storage and data size.

- Simple data preparation for modeling with any framework.
- No operation burden to build data intensive applications.
- Share and collaborate live data across company's ecosystem

2. Snowflake – Data Architecture

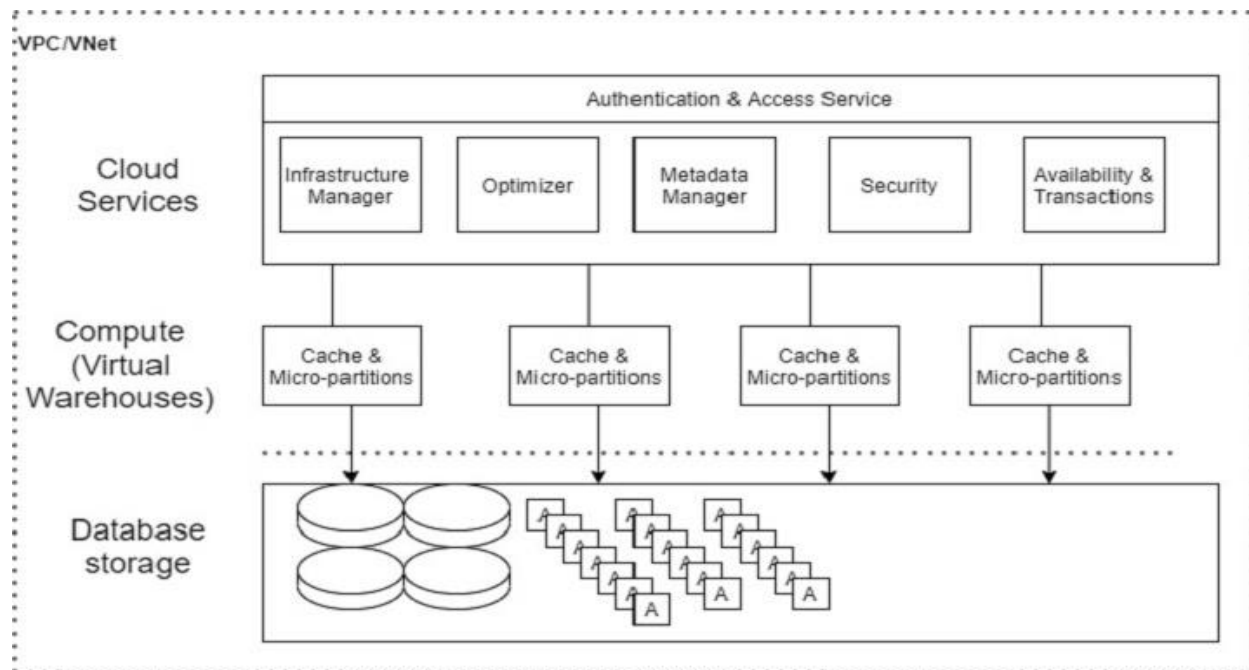
Snowflake data architecture re-invents a new SQL query engine. It is designed for the cloud only. Snowflake doesn't utilize or built on top of any existing database technology. It doesn't even use big data software platforms like Hadoop. Snowflake provides all functionalities of an analytical database plus numbers of additional unique features and capabilities to users.

Snowflake has central data repository for storage of structured and semi-structured data. These data can be accessed from all available compute nodes in the Snowflake platform. It uses virtual warehouse as compute environment for processing the queries. While processing queries, it utilizes multi-cluster, micro-partitioning and advanced cache concepts. Snowflake's cloud services are responsible to provide end to end solution to the user like logging validation of user to result of select queries.

Snowflake's data architecture has **three main layers**:

- Database Storage
- Query Processing
- Cloud Services

Following is the **data architecture** diagram of Snowflake:



Database Storage

Snowflake supports Amazon S3, Azure and Google Cloud to load data into Snowflake using file system. User should upload a file (.csv, .txt, .xlsx etc.) into the cloud and after they create a connection in Snowflake to bring the data. Data size is unlimited, but file size is up to 5GB as per cloud services. Once data is loaded into Snowflake, it utilizes its internal optimization and compression techniques to store the data into central repository as columnar format. The central repository is based on cloud where data stores.

Snowflake owns responsibilities to all aspects of data management like how data is stored using automatic clustering of data, organization and structure of data, compression technique by keeping data into many micro-partitions, metadata, statistics and many more. Snowflake stores data as data objects and users can't see or access them directly. Users can access these data through SQL queries either in Snowflake's UI or using programming language like Java, Python, PHP, Ruby etc.

Query Processing

Query execution is a part of processing layer or compute layer. To process a query, Snowflake requires compute environment, known as "Virtual Warehouse" in Snowflake's world. Virtual warehouse is a compute cluster. A virtual warehouse consists of CPU, Memory and temporary storage system so that it could perform SQL execution and DML (Data Manipulation Language) operations.

- SQL SELECT executions
- Updating of data using Update, Insert, Update
- Loading data into tables using COPY INTO <tables>
- Unloading data from tables using COPY INTO <locations>

However, the number of servers depends on size of virtual warehouses. For example, XSmall warehouse has 1 Server per cluster, while a Small Warehouse has 2 Servers per cluster and it gets double on increasing the size such as Large, XLarge, etc.

While executing a query, Snowflake analyzes the requested query and uses the latest micro-partitions and evaluates caching at different stages to increase performance and decrease the time for bringing the data. Decrease the time means less credit is used of a user.

Cloud Services

Cloud Service is the 'Brain' of the Snowflake. It coordinates and manages activities across Snowflake. It brings all components of Snowflake together to process user requests from logging validation to deliver query's response.

The following services are managed at this layer:

- It is the centralized management for all storage.
- It manages the compute environments to work with storage.
- It is responsible for upgrades, updates, patching and configuration of Snowflake at cloud.
- It performs cost-based optimizers on SQL queries.
- It gathers statistics automatically like credit used, storage capacity utilization
- Security like Authentication, Access controls based on roles and users
- It performs encryption as well as key management services.
- It stores metadata as data is loaded into the system.

And many more...

3. Snowflake – Functional Architecture

Snowflake supports structured and semi-structured data. Snowflake organizes and structures the data automatically once data loading is completed. While storing the data, Snowflake divides it on his intelligence and saves into different micro-partitions. Even Snowflake stores data into different clusters.


At functional level, to access data from Snowflake, the following components are required:

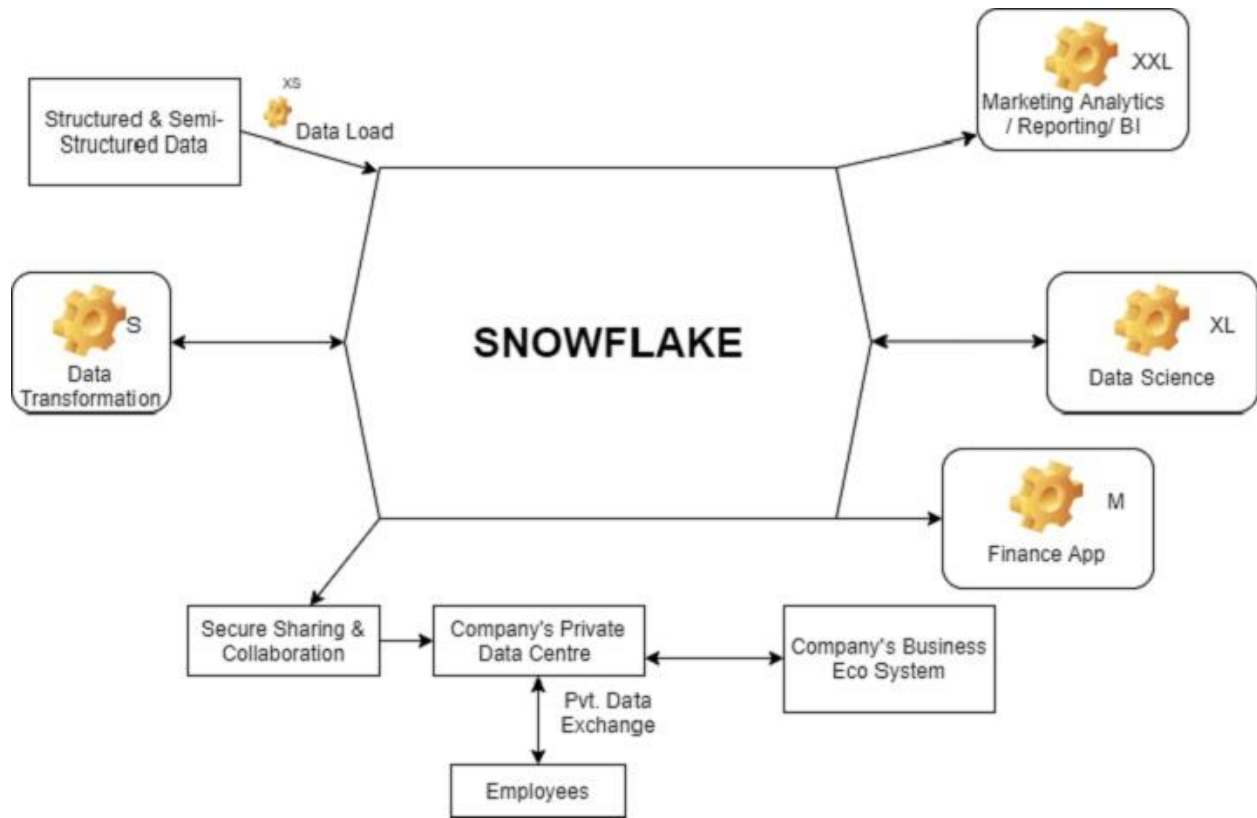
- Choose proper roles after logging
- Virtual Warehouse known as Warehouse in Snowflake to perform any activity
- Database Schema
- Database
- Tables and columns

Snowflake provides the following high-level analytics functionalities:

- Data Transformation
- Supports for Business Application
- Business Analytics/Reporting/BI
- Data Science
- Data Sharing to other data systems
- Data Cloning

The following diagram shows the functional architecture of Snowflake:

The symbol of "settings" as  in each block can be referred as Warehouse and XS, XXL, XL, L, S as sizes of warehouse requires to perform different operations. Based on requirement and usage, the size of a warehouse can be increased or decreased; even it can be converted from single cluster to multi-clusters.



4. Snowflake – How to Access

Snowflake is a licensed data platform. It uses the concept of credit to charge to clients. However, it provides 30-days free trial with \$400 credits for learning purpose.

Use the following steps to get a free access to Snowflake for 30 days:





- Open the URL "<http://www.snowflake.com>" and click "START FOR FREE" at the right corner of the page.



- It navigates to the registration page where the user needs to provide details such as First Name, Last Name, Email, Company, and Country. After filling the form, click the "CONTINUE" button.

START YOUR 30-DAY FREE TRIAL

- Gain immediate access to the Data Cloud
- Enable your most critical data workloads
- Scale instantly, elastically, and near-infinately across public clouds
- Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized

Start your 30-day free Snowflake trial which includes \$400 worth of free usage


Choose your Snowflake edition*

Standard
A strong balance between features, level of support, and cost.


Enterprise
Standard plus 90-day [time travel](#), [multi-cluster warehouses](#), and [materialized views](#).

Business Critical
Enterprise plus enhanced security, data protection, and database failover/fallback.


Choose your cloud provider*



Microsoft Azure



Amazon Web Services



Google Cloud Platform

Check here to indicate that you have read and agree to the terms of the [Snowflake Self Service On Demand Terms](#).

GET STARTED

- On the next screen, it asks to choose the Snowflake edition. Select the edition based on functionalities you want to perform. For this tutorial, standard version is sufficient but for data loading using AWS S3 we require business critical edition.
- Select **Business Critical** and then click **AWS**. Select the region where your AWS is present.
- Check the Terms & Condition box and click the **Get Started** button.

The following screenshot demonstrates the above steps:

- You will get a message that Account creation is in progress and an email is sent to your address as below screen.

- Check your email's inbox. Once you receive the email from Snowflake, usually within 2-3 mins, click the "Click to Activate" button.
- It will navigate to Snowflake's page where user requires to set up username and password. This credential will be used to login into Snowflake.
- An URL will be provided in your email something like: "https://ABC12345.us-east-1.snowflakecomputing.com/console/login". It is a user-specific URL to access Snowflake at cloud. Whenever you want to work in Snowflake, use the individual URL and login.

5. Snowflake – Editions

Snowflake provides four different editions based on users'/company's requirement.

- Standard
- Enterprise
- Business Critical
- Virtual Private Snowflake (VPS)

Standard Edition

It is the basic version of Snowflake. It provides the following features in this edition:

- Supports complete SQL Data Warehouse
- Secure Data Sharing
- Premier support for 24×365
- Time travel of 1 day
- Data Encryption
- Dedicated virtual warehouses
- Federated Authentication
- Database Replication
- External Functions support
- Snowsight
- Supporting user's own data exchange creation
- Data Marketplace Access

Enterprise Edition

It is Standard+ edition i.e. all features of Standard edition plus following additional features:

- Multi-cluster warehouses
- Time travel up to 90 days
- Encryption key changes annually
- Materialized Views
- Search Optimization Services
- Dynamic Data Masking
- External Data Tokenization

Business Critical Edition

It is Enterprise+ edition, i.e., all features of Enterprise and Standard edition plus following additional features:

- HIPPA support
- PCI compliance
- Data Encryption everywhere
- AWS Private Link Support
- Azure Private Link Support
- Database Failover and fallback

Virtual Private Snowflake (VPS) Edition

It is Business Critical+ edition and the most advanced version. It supports all the offerings of Snowflake.

- Customer dedicated virtual servers where the encryption key is in memory.
- Customer dedicated metadata store.

=====

End of ebook preview

If you liked what you saw...

Buy it from our store @ <https://store.tutorialspoint.com>